NEURAL NETS FOR DEFECT RECOGNITION ON MASKS AND INTEGRATED CIRCUITS : FIRST RESULTS

Hartmut Surmann, Benhür Kiziloglu, Ulrich Rückert & Karl Goser

University of Dortmund - Dortmund, Germany

Abstract : *The first results of applying artificial neural networks to defect detection on masks and integrated circuits will be presented. The use of neural nets opens up the possibility of developing rapid and problem-flexible inspection Systems. In this publication two different models of neural nets - the Kohonen map and a feedforward-net using the backpropagation learning algorithm - are discussed with regard to their applicability in pattern recognition for inspection Systems by means of Simulation results.*

Keywords : *self-organising feature map, backpropagation, automatic inspection System, defect recognition.*

The rapid evolution in microelectronics leads to integrated circuits with minimal structure sizes in the sub-µm-range. Such fine geometrical structures require complex and error free masks for patterning steps during the production process . Mask defects can generate local defects on the chips and can lead to functional failures or to a reduced lifetime of the devices.

For reaching a satisfactory yield, it is necessary to inspect the photomasks before using them in the production process. After fabrication an inspection of the completed semiconducting devices is essential for keeping a high quality Standard. In order to cope with decreasing structure sizes and increasing complexity of patterns, automatic inspection Systems are called for with advantages like computing speed, confident failure detection and economy.

Classical inspection Systems which are used in present fabrication processes are based on comparing one chip pattern to another or to the design data base [1]. Both methods demand for substantial computing resources in combination with the problem of a very accurate optical and mechanical adjustment. Another disadvantage is the missing classification of the defects: Differences are recognized indeed, but will not be interpreted as failures or noise.

A rather new approach is the use of artificial neural networks for defect detection on masks. Advantages like high speed (because of the massively parallel processing), adaptability on special problems, fault tolerance, and cost reduction are expected. In the following first results of accomplished research studies which point out the practical usefulness of two neural network models in the area of static image recognition (specific application: detection of mask faults) are presented and discussed.

1. Detection of mask defects

1.1 Recordlng technlques and occurlng problems

Defect recognition on masks and integrated circuits requires the recording of mask and circuit data using light-optical or electron-microscopic techniques (electron-beam scanning, X-ray scanning). All techniques use a kind of beam (light, electrons) to scan the given picture spot by spot and record the measured intensity values. The resolution of the scanning System depends on the wavelength of the used radiation: The highest resolution working with light-optical techniques is obtained by using hard X-rays. A typical scanning result is shown in Figure 1.

The picture contains a portion of noise because of external failures (like strong lighting, shadows) and inaccurate recording techniques. These failures have to be intercepted by the inspection System.

The main problem is to differ between real failures and tolerable deviations. The reduction of structure sizes in microelectronic devices leads to a necessary detection of mask and circuit defects in the sub-µm ränge.

1.2 **Occuring** failure classes

As shown in Figure 2, the possible types of failures can be devided into four main failure classes: corner, edge, bridge, pinhole/-spot. The reliability and

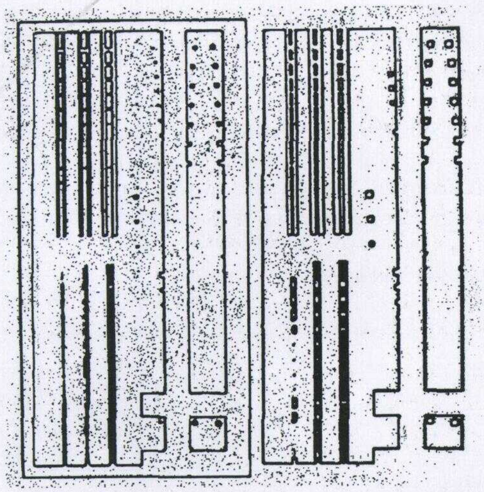
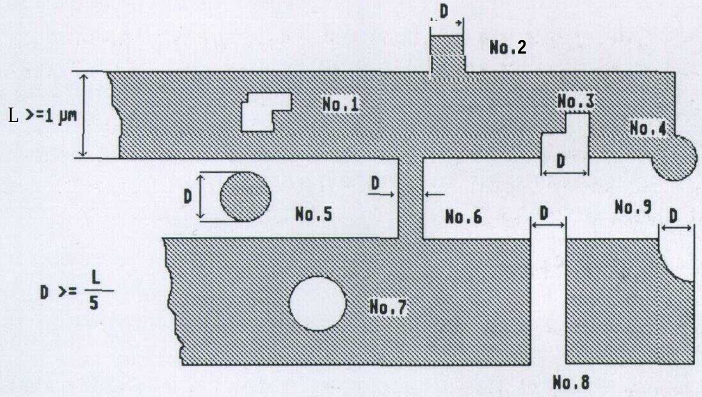


Figure 1: Half-tonepicture: 512 x 512 pixel, resolution 8 bit.



Failure classes:

Pinhole, Pinspot	Pattern No.1, S, 7
Link-Defect	Pattern No.6, 8
Edge-Defect	Pattern No.2, 3
Corner-Defect	Pattern No.4, 9

Significant data!

nininal Structure Length:	L (l µm)
nininal detectable Error-Size:	D= L/ 5 (200 m)
Resolution:	512 pixel/ 8,1 mm S pixel/ l um
Size of the Mask surface:	4 cm^2= 10^{10}pixel

Figure 2: Failure classes and significant mask data

functionality of integrated circuits depends on this failures which can eff'ect cut-offs, short circuits and reduction or enlargement of circuit structures. Figure 2 also shows some significant data of the inspected mask: The minimal detectable failure size is 0.2 um, the minimal structure size is l µm. This leads to a resolution of S pixel/µm. The resolution of the inspection System must be more than 4 or 5 times higher than the minimal structure size. By using present light-optical techniques it is possible to reach a resolution of 100 nm, what corresponds to a minimal structure size of 0.4 µm.

1.3 Conventional techniques

The most common technique deals with comparing of one mask pattern to another spot by spot. The result is only a list of differences but not a classification of the failures. This classification has to be done afterwards by the user himself. The main disadvantage of this technique is the required high adjusting accuracy. Increasing reduction of structure sizes intensifies this problem.

Another classical inspection technique is based on the verification of the design rules. Here, violations will be interpreted as possible failures. In this case the main disadvantage is given by the fact that there might be failures which are not design rule violations; these failures will not be detected. On the other hand, this technique allows a higher adjusting tolerance.

Both described techniques demand powerful computing resources for acceptable performance with regard to processing time [1].

1.4 Expected advantages of using neural networks

By using neural networks the following advantages are expected in opposite to conventional techniques:

- high performance by parallel data processing
- flexibility of the inspection Systems by learning capabilities
- insensibility against disturbances (like noise, e.g.)
- high data reduction and good classification capabilities

Parallel data processing leads to real time pattern recognition which is neccessary for powerful inspection Systems. In addition to this, neural nets offer a higher flexibility because of their ability to learn and therefore to detect new patterns. This property is important especially for smaller structure sizes.

2. Feed-forward Networks and Backpropagation

The most common used neural network architecture is a feed-forward net (FFN) **using** the well known backpropagation learning algorithm [2]. Therefore, we tried to use this architecture and learning algorithm for defect recognition as well. The general working scheme is summarized in Figure 3. After digitalizing the original mask data (Figure 1) small 8x8 pixel Windows are used as input data to our feed-forward net. The network was trained with some selected input patterns and the corresponding desired output patterns. We used different coding schemes for the Output patterns, one of them is shown in Table 1. For example, this coding scheme incorporates structual information (see part b) in Table 1) in the desired Output pattern.

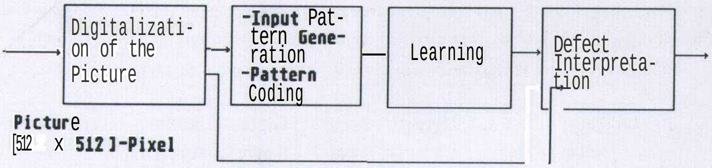
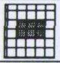
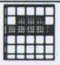
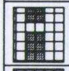
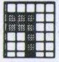


Figure 3: Working process of the feed-forward net

Failure Class	Coding (Code 1)
 Pinhole (Pinspot)	x0001
 Edge (Edge˘)	x0010
 Link (Link$^{\times}$)	x0100
 Corner (Corner$^{\times}$)	x1000

(a)

Lokal Characte-ristik	Coding (Code 2)
m	00000 00001
H	00000 00010
m	00000 00100
H	00000 01000
E	00000 10000
E	00001 00000
[Z]	00010 00000
[Z]	00100 00000
[ς]	01000 00000

(b)

Table 1: Coding technique: x=1 for faulty pattern, x=0 for correct pattern

Before simulations we have to fix a couple of System parameters. First of all the number of layers and the number of processing elments in each layer have to be selected. So far we actually use a 3 layer and a 4 layer configuration and variable number of processing elements per layer.

Then the parameters concerning the backpropagation learning algorithm have to be determined, especially the learning step width ε and the amount of weight change per learning step a [2]. Last but not least the training set and the number of training vectors have to be fixed.

Network Topology	Input-Units (Layer1)	Hidden-Units (Layer2)	Hidden-Units (Layer3)	Output-Units (Lager4)	Lear-ning Steps	End-Error	Trai-ning Time
3-Layer Net	S4	38	-	10	2789	0.0168	10min
4-Layer Net	64	30	15	10	568	0.0456	140sec

Table 2: 3-layer net and 4-layer net;
α=0.7, ε=0.8, number of patterns = 5

Concerning the number of layers our simulations have shown no remarkable influence on System behaviour except an increasing number of necessary training Steps for the 3 layer network in Order to get the same recognition performance (Table 2). The number of processing elements per layer has in contrary a perceivable impact on the convergence of the learning phase. According to that an increasing number of processing elements do not necessarily lead to a better System performance (Table 3).

Network Topology 4-Layer Net	Input-Units (Layer1)	Hidden-Units (Lager2)	Hidden-Units (Layer3)	Output-Units (Lager4)	Learning Steps	End-error	Training Time
Net-A	100	30	15	15	261	0.039	2 min
Net-B	121	30	15	15	264	0.043	2 min
Net-C	144	3B	15	15	4000	3.2	30min

Table 3: Number of patterns = 5, α=0.7, ε=0.8

The learning Parameter ε (learning step width) has a substantial influence on the convergence behaviour during learning as can be seen from Tables 4 and 5. A suitable choice of E leads to a faster convergence. The parameter a influences the convergence behaviour as well (Tables 4 and 5). The maximal number of patterns successfully learned by our used network topologies was 25 and especially independent of the choice of the learning set. Further Simulation results of the learning behaviour for different parameter choices and network topologies are given in Figures 3-5. All the simulations have be done on a HP 9000 (Unix, C-language, MC 68030, 50MHz, 24 MB RAM).

Network Topology 4-Lager Net	α	E	Number of Patterns	Learning Steps	End-error	Training Tine
Net-B2	0.7	0.8	25	6000	17.539	$3\frac{1}{2}$h
Net-B3	0.7	0.4	25	taee	33.297	$3\frac{1}{2}$ h
Net-B4	0.8	0.4	25	6000	31.359	$3\frac{1}{2}$ h

Table 4: Network topology as net-B1 in Table 3

Network Topology 4-Lager Net	α	ε	Nunber of Patterns	Learning Steps	End-error	Training Tine
Net-D1	0.7	8.8	25	6000	35.0435	$3\frac{1}{2}$ h
Net-P2	0.7	0.4	25	6000	25.6064	$3\frac{1}{2}$ h
Net-D3	0.8	0.4	25	3179	1.3491	2 h
Net-D4	8.7	0.4	10	362	0.0477	3 min
Net-D5	0.7	B.B	1B	1BBB	0.5120	13nin

Table 5: Network topology [121 3O 15 10l

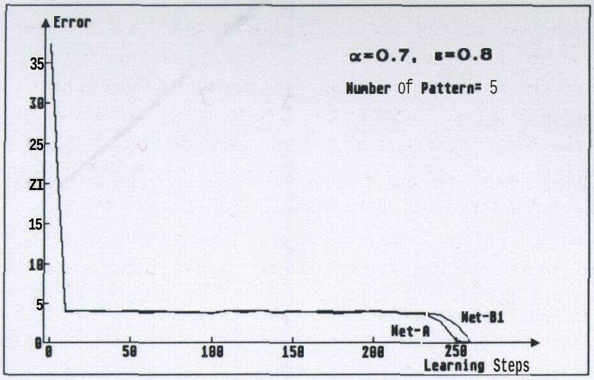


Figure 3: Simulation results for the net-A and net-B1

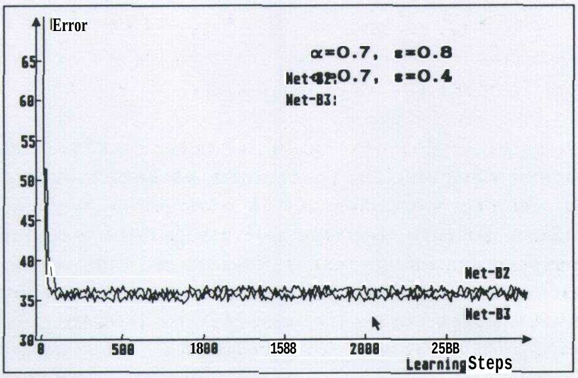


Figure 4: Simulation results for the net-B2 and net-B3

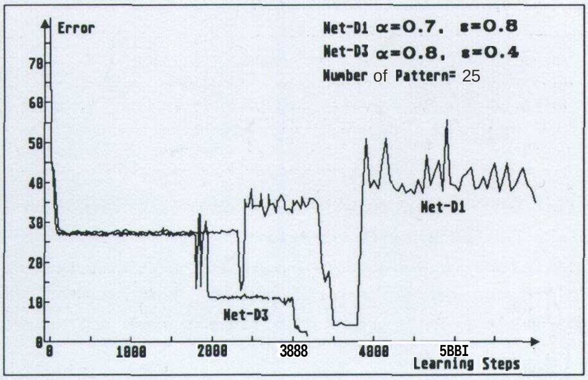


Figure 5: Simulation results for net-D1 and net-D3

In summary, our results confirm the experience of many other research groups dealing with the backpropagation learning rule. The main problem is a proper choice of the network topology and the System Parameters for a given application. For our indendet application (defect recognition on masks and integrated circuits) we have not succeeded in solving this Problem so far. Hence, our first result are not very promising.

3 The self-organizing feature map

The self-organizing feature map consists of a two-dimensional array of identical processing units. Each of this processing units stores a single vector, where each vector component is connected to the corresponding component of an external input vector. As a learning aigorithm we use the Kohonen aigorithm [3],[4].

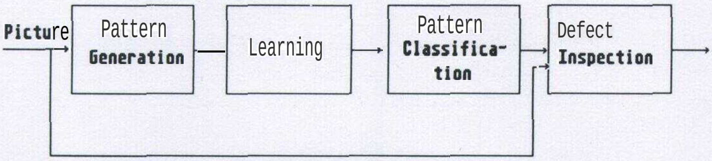


Figure 8: *Working process* for the feature map

The working process can also be organized in four Segments as seen in Figure 8. In the first Step we filter out 16 x 16 pixels with 8 Bit resolution. As the self-organizing feature map can not assign the learning pattern automatically to the error classes, an expert has to do this work after training. During the classification process, he choose different parts of the feature map and assign these parts to the failure classes in Table 6.

Code	Classification
0	no Error
1	Pinhole-Error
2	Pinspot-Error
3	Edge-Error
4	Corner-Error
5	Link-Error

Table 6: Failure classes

One problem of the self-organizing feature map is to find an optimal number of processing units. The number is correlated with the number and differences in the learning patterns. If we use not enough processing units, the feature map could not find some characteristics during the inspection process. If we have to much processing units then the time grows up exponentialy during the training and inspection process.

In the same way we have simulated the FFN, we have performed several simulations with different numbers of processing units (144, 196, 1024) and patterns. Figure 10 denotes one good result. It shows for every processing unit in the feature map the adapted 16x16 pattern mask. It can be seen that similiar input patterns (in sence of the euklidian norm) are mapped to processing units in the same region of the map. In Table 7 the learning and inspection times for our Simulation program can be seen (HP 9000).

Network Topology	h0, h1, bB, b1, ni, nj, Inputunltflnz	Number of Patterns	Learning Steps	Training Time	Error Evaluation-Time
Net-KR	0.9, 0.02, 9.0, 1.00, 14, 14, 256	3«	30	S $\frac{1}{2}$ h	—
Net-KB	0.9, 0.02, 8.0, 1.00, 14, 14, 256	»S	30	53 min	25 min
Net-KB	0.9, 0.02, 8.0, 1.B8, 16, 16, ZS6	Z5S	30	88 min	—

Table 7: Learning and inspection times
hO, b0: learning step and radius size at the beginning
hl, bl: learning step and radius size at the end
ni, nj: number of processing units in x and y direction [4]

During the inspection process we generate 16 x 16 pixel input patterns from the orginal mask pictures. The user can choose an offset in x and y direction in our Simulation program. After the generation the input pattern is classified by the feature map.

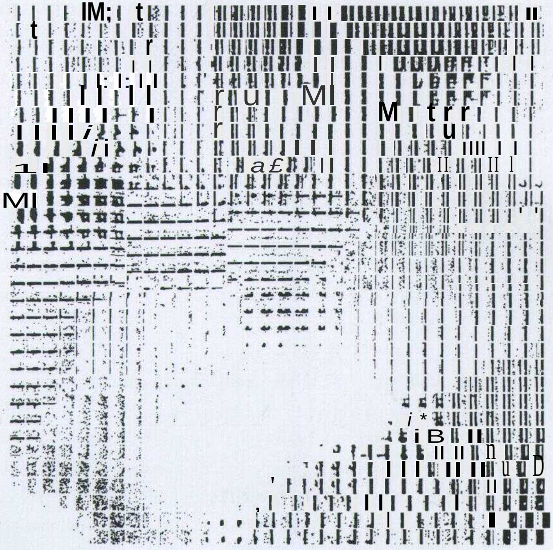


Figure 10: Weight vectors of the self-organizing feature map after learning. Each weight vector represents a small 16 x 16 picture.

In our examination we could successful inspect masks automaticaly with the knowledge of the feature map. Figure 11 shows a mask, which has been inspected with an offset x=y=8. If we use smaller offsets (e.g. x=y=4) we detect more errors, but the inspection time grows up. For a real-time inspection System it is necessary to have special neural network hardware [5], [6].

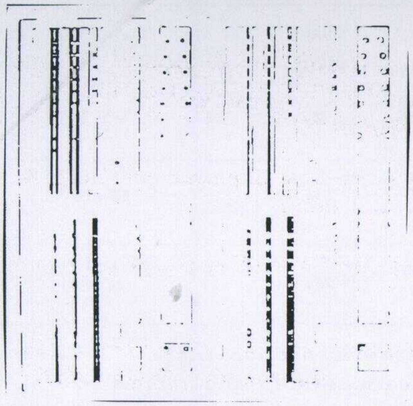


Figure 11: Result of an inspection process. Offset x=y=8. Errors are marked with a frame.

4. Conclusion

In this paper two neural network models - the self-organizing feature map and several feed-forward nets - have been analysed with regard to their applicab ility in defect recognition on microelectronic structures.

The Simulation results concerning the feed-forward nets have shown that the training procedure will be more complicated with a rising complexity in the number of patterns and input elements. The important question now is whether the neural network model is generally unsuitable for the problem of defect recognition or the selected parameters to be adjusted (number of hidden elements, coding, . . .) prevent a satisfactory learning characteristic. Our experiences and the appeared problems are similar to those of other research groups.

On the other hand, the Simulation of Kohonen's algorithm have led to useful results which are applicable for failure diagnosis. The Interpretation of the Kohonen net have shown good generalization results so that noisy patterns have been correctly classified. In further examinations, modified learning procedures with different pattern codings as well as automatically assigning the different parts of the feature map to the error classes will be analysed. Furthermore, we have to develop a "recognition measure", so that we can say something about the recognition efficency or quality. Up to now other inspection System (see paragraph 1.3) could only show differences but they cannot classify these differences. So that neural networks can be regardet a useful addition to conventional inspection System at least.

Acknowledgements

The authors would like to thank the Siemens AG Munich especially Dr. E. Wolfgang and Dr. U. Ramacher and the Sietec GmbH & Co. OHG (Siemens-Systemtechnik, Berlin) for supplying us with mask data and Stefan RUping as well as Thomas Will for their assistance in preparing the manuscript.

Literature

[1] Sischka, D., Bisek, R.: "Detection of Defects on the Surface of Microelectronic Structures", IEEE Transaction on Electron Devices, Vol. 36, No. l, January 1989, pp 8-13.

[2] McLelland, J., Rummelhart, D.: "Explorations in parallel distributed processing." MIT Press, London 1988.

[3] Kohonen, T.: "Self-Organisation and Associative Memory", Springer Verlag, 1984.

[4] Ritter, H., Marinetz, T., Schulten, K.: "Neuroinformatik selbstorganisierender Abbildungen", Addison-Wesley-Verlag, 1989.

[5] Tryba, V., Speckmann, H, S., Goser, K.: "A Digital Hardware-Implementation of a Selforganizing Feature Map as a Neural coprocessor to a Von - Neumann Computer", 2st Int. Workshop on Microelectronics for Neural Networks June 1990, pp 177 - 186.

[6] Ramacher, U., RUckert, U.: "VLSI Design of Neural Networks" Kluwer Academic, 1991.